

Seguimiento de la campaña “all you need is Ecuador” en Twitter

. *Revista Publicando*, 2(4). 2015, 2-10. ISSN 1390-9304

Seguimiento de la campaña “all you need is Ecuador” en Twitter

Jorge Alberto González Alonso¹, José Francisco Taipe Yáñez², Johny Fabián

Pazmiño Linzán³, Yudeisy Pérez González⁴

1 RML Consultores, proyectos@rmlconsultores.com

2 Universidad Central del Ecuador, jtaipe@uce.edu.ec

3 Universidad Central del Ecuador, globalserviceelge@gmail.com

4. Centro de Estudios Martianos, yudeisy200731@gmail.com

RESUMEN

Empleando la herramienta de minería de datos KNIME se obtuvo una relación de los mensajes enviados, durante aproximadamente cinco días en Twitter y que empleaban el hashtag #allyouneedisecuador.

La agrupación por clústeres realizada empleando el algoritmo de k-medias permitió clasificar estos mensajes en cinco grupos de acuerdo con los indicadores: reenvío y mensajes marcados como favoritos.

Los resultados son alentadores en cuanto a la herramienta empleada: KNIME, para obtener los mensajes en Twitter y en cuanto a la posibilidad de agrupamiento en clústeres.

El trabajo realizado ofrece al analista de campaña diferentes posibilidades de filtrar los resultados de acuerdo con reenvíos y analizar las características de estos mensajes y usuarios, por ejemplo agencias de viajes con más impacto, o menos en relación con la campaña analizada.

Palabras clave: minería de datos, knime, #allyouneedisecuador, twitter



Seguimiento de la campaña “all you need is Ecuador” en Twitter

. *Revista Publicando*, 2(4). 2015, 2-10. ISSN 1390-9304

Title of the paper in English language

ABSTRACT

Using a data mining tool such as KNIME messages sent over a period of five days, on Twitter and using the hashtag #allyouneedisecuador, could be obtained

The messages were classified in clusters using the k-mean algorithm and it was possible to classify these messages on five groups according to the indicators and forwarding messages and favorites.

The results are encouraging in terms of both the tool used to collect the Twitter messages: KNIME, as regarding the possibility of clustering.

The work allows to offer the analyst of campaign different possibilities to filter the results according to forwards and favorites and analyze the characteristics of these messages and users, such as travel agencies, with more impact, or less in relation to the analyzed campaign.

Keywords: data mining, knime, #allyouneedisecuador, twitter



Seguimiento de la campaña “all you need is Ecuador” en Twitter

. *Revista Publicando*, 2(4). 2015, 2-10. ISSN 1390-9304

1. INTRODUCCIÓN

Desde hace ya varios años las Redes Sociales y en particular Twitter son reconocidas como una fuente que permite conocer sentimientos y opiniones (Pak & Paroubek, 2010) en variadas situaciones (Burnap et al., 2014). El denominado “análisis de sentimientos” se ha convertido así en una línea investigativa (Agarwal, Xie, Vovsha, Rambow, & Passonneau, 2011) con sus pros y contras a favor de lo confiable de sus resultados (Kouloumpis, Wilson, & Moore, 2011).

El hecho cierto es que el “microblogging” es cada día más popular entre los usuarios de Internet y la minería de datos una de las técnicas aplicadas para conocer las opiniones que se transmiten de esa forma (Pang & Lee, 2008).

En el caso particular de este reporte investigativo se enfoca en seguir una campaña publicitaria a partir de un denominado “hashtag”, es decir de un marcador dado que es el que caracteriza los mensajes transmitidos en relación con la misma empleando la plataforma Twitter.

La campaña a seguir en Twitter se conoce como “**All you need is Ecuador**” y puede caracterizarse como uno de los proyectos más grandes de Marketing en América Latina impulsado por el Ministerio de Turismo de Ecuador (Turismo, 2015) para la captación de turistas de EEUU y lanzado en un spot televisivo en el Super Bowl 2015. Para seguir el flujo de Tweets se empleará el hashtag de la campaña: #allyouneedisecuador y como herramienta el KNIME (Berthold et al., 2007) utilizado para la minería de datos y el análisis de sentimientos (Minanovic, Gabelica, & Krstic, 2014).

Este reporte de investigación se inscribe dentro de un proyecto más general relacionado con el empleo de herramientas de minería de datos para estudiar el marketing en redes sociales y se propuso como objetivo la clasificación de los mensajes en Twitter que utilizaran #allyouneedisecuador durante un período de tiempo dado empleando el KNIME para determinar la posible utilidad de la información obtenida para el seguimiento de la campaña publicitaria que utiliza el marcador mencionado.

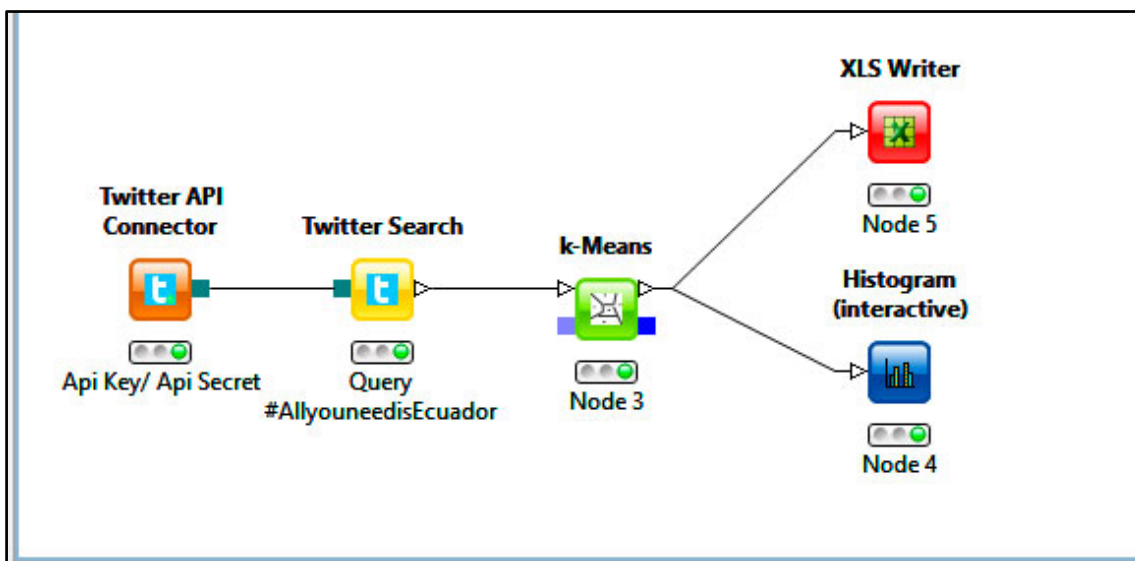


“ ”

2. METODOS

La investigación realizada se desarrolló con el empleo de los siguientes procesos que se esquematizan en la Figura 1 a continuación y que pueden implementarse en KNIME (Berthold et al., 2007);

Figura 1. Esquema de procesos en KNIME



Como se observa el diagrama presentado comprende los sub procesos siguientes:

1. Conexión con la API de Twitter. Para ello se creó una Aplicación (App) en Twitter, la documentación y posibilidades están documentadas por Twitter en (Twitter, 2015). La clave de esa App (Api Key) se introduce en el Conector junto a la clave secreta (Api Secret) enviada por Twitter. Es decir la única función del primer subproceso es simplemente obtener la comunicación autorizada con la Plataforma Twitter.
2. Búsqueda en Twitter. En este módulo si se puede ya realizar la búsqueda para el término en este caso: #allyouneedisecuador. Se puede configurar igualmente la búsqueda para el tipo de mensaje, por ejemplo más recientes, populares o para ambos casos siendo esta última opción la seleccionada. Este subproceso da en su



Seguimiento de la campaña “all you need is Ecuador” en Twitter

. *Revista Publicando*, 2(4). 2015, 2-10. ISSN 1390-9304

salida 2500 mensajes, enviados en un plazo de cinco días. La Tabla presenta los campos siguientes: usuario, tweet, día y hora de envío, si ha sido favorito de otro usuario, si es un reenvío de otro y de serlo de cual usuario.

3. El subproceso siguiente agrupa en clústeres toda la información, empleando el algoritmo de k.medias, en este caso se seleccionó a partir de pruebas previas el valor $k=5$
4. Los dos subprocesos siguientes toman la información obtenida del proceso de clusterización anterior y son:
 - a. Escribir los resultados de la clusterización en un fichero en Excel
 - b. Preparar un histograma interactivo que visualiza el número de mensajes por clúster y a la vez la cantidad de favoritos y reenvíos

Hay que destacar que en la búsqueda en Twitter se puede configurar también la carga de imágenes, es decir se puede obtener no sólo el mensaje sino la imagen utilizada, en este caso el tiempo de búsqueda es solo mayor.

3. RESULTADOS

La Base de Datos obtenida directamente en Excel, permite una rápida clasificación empleando filtros para los campos:

- Usuario
- Mensaje (Tweet)
- Día, hora
- Favorito
- Reenvio (Rtweeted)
- Reenvio de (Retweet from)

Sólo a modo de ejemplo mostramos los mensajes que fueron marcados por más de 10 usuarios



Seguimiento de la campaña “all you need is Ecuador” en Twitter

. Revista Publicando, 2(4). 2015, 2-10. ISSN 1390-9304

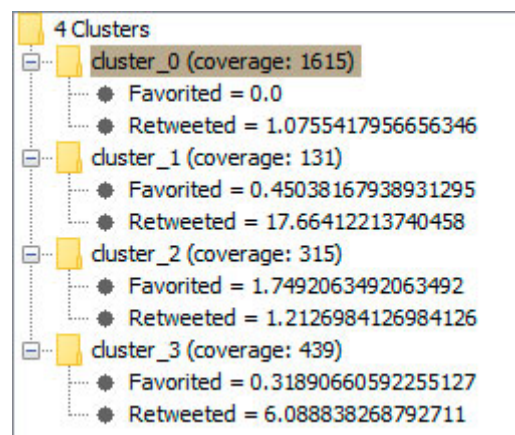
Tabla 1. Ejemplo de seguimiento de mensajes con más impacto

User	Tweet	Time	Favorited	Rtweeted
makecuador	#Loja Un hermoso atardecer en el cantón Catamayo. #AllYouNeedIsEcuador @makecuador http://t.co/GyOCjMTEjA via @primereporte	2015-08-02 16:59:08	13	23
CGEEUUPatricia	Increible! Jugo de cacao sabe parecido a jugo de guanabana y es delicioso! #AllYouNeedIsEcuador @EmbajadaEcuUSA http://t.co/KQRRJCR698	2015-07-30 19:49:01	12	16
Sandynarango	Más de 20 nominaciones 🏆 y más de #1000Razones para decirle al mundo #AllYouNeedIsEcuador http://t.co/y1J14m3hrw http://t.co/Jh55ZlvsAh	2015-08-04 08:23:02	11	18
Makecuador	Guayaquil desde los#444Escalones #allyouneedisecuador @makecuador @AllYouNeedisEC 😊😊❤️ http://t.co/0fRrDeTtCT via @LadyDi096	2015-08-03 18:37:04	10	9

Esta Tabla ejemplifica claramente la posibilidad de seguimiento diario de mensajes de mayor impacto por personal sin un nivel de calificación elevado.

La clasificación por clúster permitió agrupar el total de mensaje de acuerdo con los centroides siguientes por clúster:

Fig. 2. Resumen de datos de agrupamiento para los centroides de cada clúster





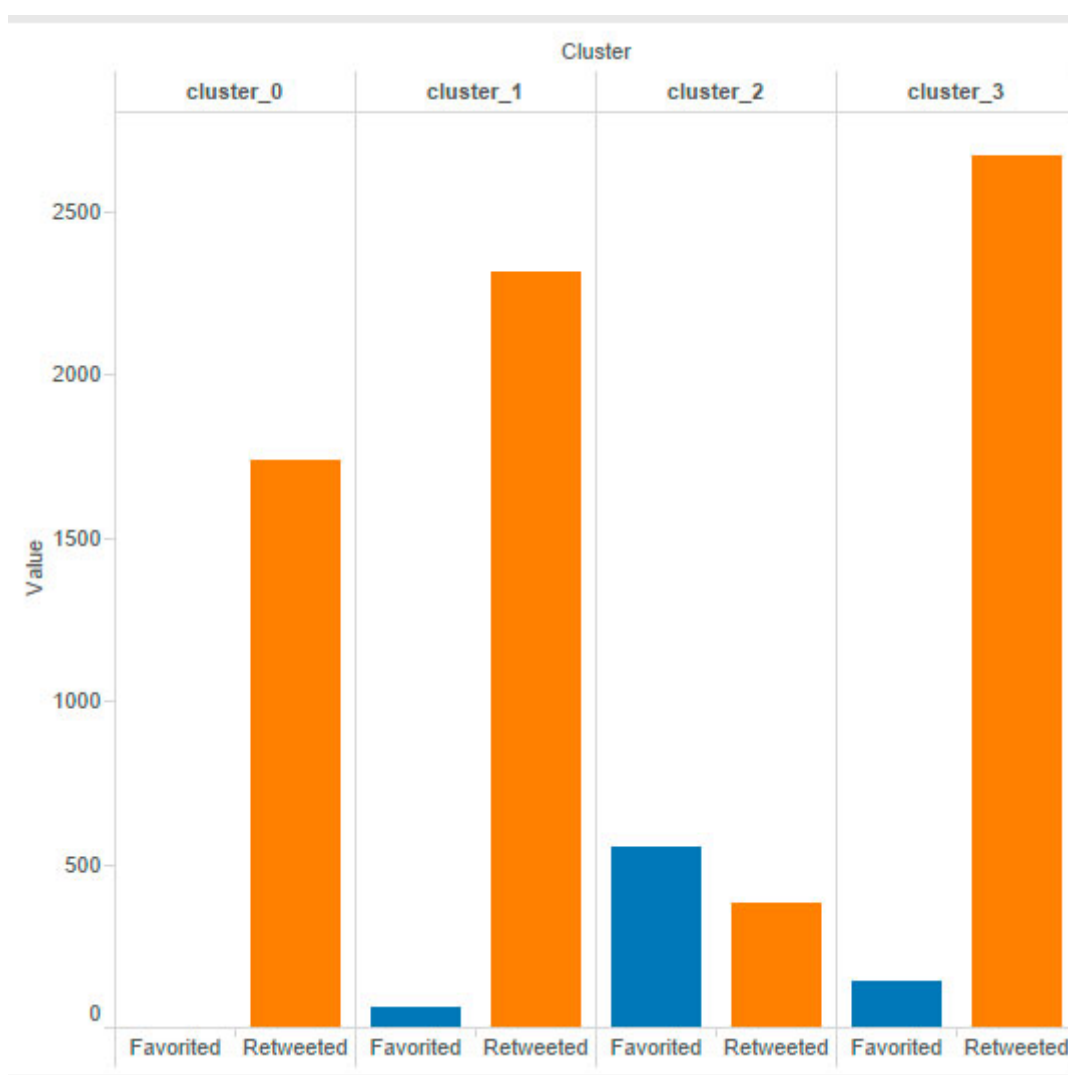
Seguimiento de la campaña “all you need is Ecuador” en Twitter

. *Revista Publicando*, 2(4). 2015, 2-10. ISSN 1390-9304

Como se observa el mayor número de mensajes se agrupa en el clúster 0, un total de 1615 mensajes que no han sido marcados como favoritos por otros usuarios y que también como promedio sólo se han reenviado una vez.

La figura a continuación ejemplifica la cantidad de mensajes reenviados o marcados como favoritos para cada clúster:

Figura 3. Agrupamiento por clústeres de los mensajes reenviados o marcados como favoritos



Desde la apreciación del analista de campaña queda claro que se puede realizar un análisis detallado de cuáles son los mensajes que más o menos impacto generan y por qué.



Seguimiento de la campaña “all you need is Ecuador” en Twitter

. *Revista Publicando*, 2(4). 2015, 2-10. ISSN 1390-9304

Una forma simple es por ejemplo filtrar los resultados de acuerdo con reenvíos y analizar las características de estos mensajes. No es objetivo de este reporte precisar ese análisis pero si podemos apuntar que los resultados obtenidos, nos permiten por ejemplo conocer las agencias de viaje que más o menos impacto, tienen en relación con la campaña al utilizar el hashtag #allyouneedisecuador, o considerar otros posibles hashtags de apoyo.

4. CONCLUSIONES

Empleando KNIME se pudo seguir en detalle los mensajes enviados, durante un período aproximado de cinco días, por Twitter y que empleaban el hashtag #allyouneedisecuador.

La agrupación por clústeres realizada permitió clasificar estos mensajes en cinco grupos de acuerdo con los indicadores: reenvío y marcados como favoritos.

Los resultados son alentadores en cuanto a la herramienta empleada: KNIME, para obtener los mensajes en Twitter y en relación con la posibilidad de agrupamiento en clústeres.

El trabajo realizado permite ofrecer al analista de campaña diferentes posibilidades de filtrar los resultados de acuerdo con reenvíos y analizar las características de estos mensajes y usuarios, por ejemplo agencias de viajes con o menos más impacto, en relación con la campaña analizada.



5. REFERENCIAS BIBLIOGRÁFICAS

- Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). *Sentiment analysis of twitter data*. Paper presented at the Proceedings of the Workshop on Languages in Social Media.
- Berthold, M. R., Cebron, N., Dil, F., Gabriel, T. R., Kotter, T., Meinl, T., . . . Wiswedel, B. (2007). *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*: Springer.
- Burnap, P., Williams, M. L., Sloan, L., Rana, O., Housley, W., Edwards, A., . . . Voss, A. (2014). Tweeting the terror: modelling the social media reaction to the Woolwich terrorist attack. *Social Network Analysis and Mining*, 4(1), 1-14.
- Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter sentiment analysis: The good the bad and the omg! *ICWSM*, 11, 538-541.
- Minanovic, A., Gabelica, H., & Krstic, Z. (2014). *Big data and sentiment analysis using KNIME: Online reviews vs. social media*. Paper presented at the Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on.
- Pak, A., & Paroubek, P. (2010). *Twitter as a Corpus for Sentiment Analysis and Opinion Mining*. Paper presented at the LREC.
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2), 1-135.
- Turismo, M. d. (2015). Ministerio de Turismo. *Comunicamos*. Retrieved 1 de Junio, 2015
- Twitter. (2015). Developers Documentation. from <https://dev.twitter.com/overview/documentation>